Biostatistics Research Core (BRC)
Policies for Statistical Collaboration

General Policies

- BRC project requests are reviewed and assigned to a statistician weekly.
- Upon assignment, the statistician(s) will contact the investigators to set up an initial consultation and determine scope of the work.
- Mentors for trainees and new investigators must attend the initial consultation with the statistician.
- If the project includes more than a single initial consultation with a BRC statistician, the BRC statistician will estimate time to complete the agreed upon scope of work.
- BRC administration will review the estimated hours of required statistical support and the scope of the project. If the estimated amount of BRC resources and/or personnel time to complete the project are more than the institutional resources can feasibly cover at that time, BRC administration may require funds to support the BRC's efforts on the project.
- Administration will allocate support on an individual basis. While the BRC is highly committed to providing in-kind support to trainees and new investigators, there is no guarantee of gratis statistical support for any individual project. If the investigator(s) have difficulty obtaining funds for the required statistical support, the BRC administration may assist with providing information in seeking out funds.
- You must meet with a statistician **before** collecting data and in the design phase of the research process.
- Data should be cleaned and prepared as outlined in the attached "Data Preparation for Statistical Analysis" checklist. All items must be read and completed prior to sending data to the statistician.
- Occasionally, statisticians are asked to review analyses run directly by the investigator or elsewhere. These reviews are generally treated as full analyses and require the same period of time to complete.

Timeline

- Timelines for projects may vary based on the scope of work and current queue of projects. We will develop an appropriate timeline for the scope of work after an initial consultation. In general, we advise allowing **four to six weeks** from receipt of the data for completion of analysis.
- Please also allow **four to six weeks** for help with grant applications, abstracts, and manuscript reviews.

Authorship

- As with any paper or abstract, authorship is generally expected for collaborative efforts, whenever substantive input on the design or analysis is provided. The BRC endorses the criteria recommended by the International Committee of Medical Journal Editors (ICMJE). These criteria state that authorship credit should be based on:
  - Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
  - Drafting the work or revising it critically for important intellectual content; AND
  - Final approval of the version to be published; AND
  - Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data Preparation for Statistical Analysis

DATA FILE:

- All data should be on ONE excel spreadsheet and not multiple tabs in a file. If you have multiple sheets, incorporate all information onto one sheet.
- Do not color code or have blank rows within your spreadsheet.
- Only variables needed for analysis should be included. Text fields with notes should be removed from the spreadsheet. When you hide a column in excel, it will still be read into statistical software; therefore, delete unnecessary columns.

VARIABLE NAMES:

- Row 1 should contain unique variable names and preferably not longer than 12 characters (not too long but not so short as not to be descriptive).
- Variable names should not begin with a number.
- The first patient data or observation should be entered on row 2.
- Each row will be a separate patient or observation.
- Do not include symbols in variable names. Underscores are an option, but no blanks or spaces between words.

SUBJECT/OBSERVATION IDs:

- Have the first column for variable name: **ID,** which should be a unique number for each independent study subject/observation. No names, MRNs, or other personally identifiable information (PII) should be included per HIPAA privacy rules. Ideally, start with 1 and end with the number of study subjects/observations. Excel row numbers will not work. This number will serve as a reference number when referring to a subject/observation.
- Note that if there are multiple rows or observations per subject, you should include the subject ID in each row.

CODING:

- Code all data numerically (e.g. 0=no, 1=yes, male=1, and female=0), and **include a data codebook or a data dictionary**, either in a separate sheet or word document.
- If your dataset has scale variables, please identify what the scale means. (e.g., pain score 0-10, 0=no pain, 10=severe pain).
- If a particular scale has values such as “1+”, “2+”, etc., remove the symbols (“+”) and recode such that data are only numeric (consider replacing with “1.5”, “2.5”).
- Do not code “missing”, or “not done” as “ND” or “N/A”; consider coding as “999” or “888”, or simply leave blank. Be sure to include any missing codes in your data dictionary.
- When referring to data, use variable names not excel column letters. Once the data is in the statistical package, there is no column A, B, C or X, Y, Z.

Updated: November 2017